The Security-Utility Trade-off for Iris Authentication and Eye Animation for Social Virtual Avatars

Brendan John, *Student Member, IEEE*, Sophie Jörg, Sanjeev Koppal, *Senior Member, IEEE*, and Eakta Jain

Abstract—The gaze behavior of virtual avatars is critical to social presence and perceived eye contact during social interactions in Virtual Reality. Virtual Reality headsets are being designed with integrated eye tracking to enable compelling virtual social interactions. This paper shows that the near infra-red cameras used in eye tracking capture eye images that contain iris patterns of the user. Because iris patterns are a gold standard biometric, the current technology places the user's biometric identity at risk. Our first contribution is an optical defocus based hardware solution to remove the iris biometric from the stream of eye tracking images. We characterize the performance of this solution with different internal parameters. Our second contribution is a psychophysical experiment with a same-different task that investigates the sensitivity of users to a virtual avatar's eye movements when this solution is applied. By deriving detection threshold values, our findings provide a range of defocus parameters where the change in eye movements would go unnoticed in a conversational setting. Our third contribution is a perceptual study to determine the impact of defocus parameters on the perceived eye contact, attentiveness, naturalness, and truthfulness of the avatar. Thus, if a user wishes to protect their iris biometric, our approach provides a solution that balances biometric protection while preventing their conversation partner from perceiving a difference in the user's virtual avatar. This work is the first to develop secure eye tracking configurations for VR/AR/XR applications and motivates future work in the area.

Index Terms—Security, Eye Tracking, Iris Recognition, Animated Avatars, Eye Movements

1 Introduction

Eye tracking will transform virtual and mixed reality. Major hardware companies are integrating eye trackers into head-mounted displays (HMDs) to enable applications ranging from intuitive gaze-based interfaces [63, 68, 86], foveated rendering [9, 64], and streaming optimization [24,52]. Foveated rendering is driving eye tracking within VR headsets due to the potential to both optimize resources and reduce simulator sickness [64, 78]. For social virtual reality with hyper-realistic virtual avatars [51, 53], eye tracking is required to transfer non-verbal social cues from the user to his or her conversational virtual avatar.

Because of the networked nature of social platforms and the use of cloud-based rendering techniques for VR [57], it is expected that XR devices will follow an 'always on and connected' model. Streaming eye tracking data makes it susceptible to attacks. Most critically, the iris image of the user is vulnerable. The iris image is a gold standard biometric that is used in high security applications, such as border customs [2], and is recognized as such by headset manufacturers [4]. John et al. [42] showed that typical eye tracker eye images, if stolen, could be used to biometrically identify as a user. They presented a proof of concept solution that blurred the eye image to remove the high frequency patterns that form each person's unique iris signature. They evaluated this solution for a target viewing task. However, for such a solution to be impactful, it is also necessary to determine the consequences of a security mechanism for specific applications. We focus on the application of eye tracking to animate the eyes of virtual avatars, as eyes are critical to realism and naturalness of avatars, gaze is a crucial social cue in conversations, and inadvertently altering a user's gaze may result in unintended changes in how he or she is perceived.

Our first contribution is to discuss the theoretical basis of this problem and a proposed solution, provide a novel hardware mechanism to achieve the solution, and evaluate its ability to reduce accuracy of iris authentication. Our second contribution is to determine detection thresholds for the amount of image defocus that can be applied before a difference in eye animations is perceived. Our third contribution is a study to determine how image defocus impacts perceived eye contact, attentiveness, naturalness, comfort, and truthfulness of the conversational avatar. Based on this work, it is possible to recommend to a user how to create their preferred level of security for eye tracking, and how much impact this setting will have on the perceived characteristics of their virtual avatar. More broadly, this work motivates the need to investigate the security-utility tradeoff for a wide range of XR applications and develop eye tracking configurations that prioritize security.

- *Brendan John is a PhD student at the University of Florida. E-mail: brendanjohn@ufl.edu.*
- *Dr. Sophie Jörg is an Associate Professor at Clemson University. E-mail: sjoerg@clemson.edu.*
- *Dr. Sanjeev Koppal is an Assistant Professor at the University of Florida. E-mail: sjkoppal@ece.ufl.edu.*
- *Dr. Eakta Jain is an Assistant Professor at the University of Florida. E-mail: ejain@cise.ufl.edu*

DOI: 10.1109/TVCG.2020.2973052

2 Background

Eye Tracking in Virtual Reality Current applications of eye movements in VR are driving investments in the next generation of eye tracking hardware. Applications include foveated rendering [9, 64], which optimizes computational resources in rendering by reducing resolution in the periphery, streaming algorithms that reduce the bandwidth of streamed 360° content [24, 38, 52], intuitive interfaces for navigation and predicting intent [10, 63], subtle gaze direction using luminance cues in the periphery to guide attention [31], redirected walking methods that take advantage of saccadic masking and blinks to orient the user within a limited physical space [46, 79], classifying neurodegenerative disease through eye movements [62], virtual experiences designed to improve joint attention of children with ASD [55], and modeling how users explore 360° content [73]. Eye tracking hardware in VR ranges from video-based oculography [45], electro-oculography (EOG) [13], photo-sensor oculography (PS-OG) [48, 84], and magnetic sclera coils [83]. EOG, PS-OG, and sclera coil eye trackers provide gaze estimation without imaging the eye itself, however video-based eye trackers are the most readily available solutions today. EOG and sclera coil approaches are invasive, as they require electrodes to be attached to the user's head or a magnetic contact lens to be worn by the user. PS-OG trackers are still being evaluated in terms of power usage and the ability to deploy within consumer devices, as the current implementation occludes the user's field of view [84]. Companies like Facebook, HTC, and Magic Leap have opted for a non-invasive video-based eye tracker that captures images of the eye, including the iris and

other identifiable features like eyebrows [21]. Thus, there is a need to investigate techniques that secure the iris during gaze estimation.

Eye Movements for Conversational Virtual Avatars Eye movements play an important role in non-verbal communication, and thus are critical in creating compelling social interactions with virtual avatars. For example, Steptoe et al. [77] showed that the presence of eye movements caused participants to more accurately determine if an avatar was being truthful or not when compared to an avatar without eye movements. This is important for conversational avatars that discuss sensitive information, such as medical diagnoses [82]. The animation of virtual eyes can be data-driven or generated by procedural algorithms that model the dynamics of the eye. Realistic eye animations may include characteristics such as micro-saccadic jitter, blinks, eye lid displacement, and pupil diameter [71]. Results from Duchowski et al. [23] suggest that data-driven eye animations are perceived as more natural than procedural animations. Jörg et al. [43] found that subtle variations in the amplitude of noise within data-driven eye animations influenced how natural the animations were perceived. This suggests that a small amount of spatial noise in the signal may be detected, and have a negative impact on the naturalness of eye animations. Results from Garau et al. [30] suggest that a virtual avatar rendered with naturalistic eye and head movements did not improve communication over an audio-only conversation, when the eye and head movements do not match the context of the conversation. The authors also showed that an avatar with eye movements based on the current conversation produced similar responses in attentiveness and involvement to that of a video call with a real person. This implies that while models can be used to generate natural eye movements for an avatar, they may not contain the non-verbal cues and subtleties needed to simulate a real conversation. In these cases real eye tracking data is critical. In this paper we focus on data-driven eye movements in the absence of cues like blinks, eyelid movement, or pupil dilation to isolate the influence of perceived gaze direction and dynamics of the eyeball.

Privacy & Security in Eye Tracking There is a growing concern in keeping eye movement data private and secure in both real-time applications [49], and published datasets [50]. Publicly available datasets release de-identified gaze data from individuals viewing VR videos [19], the social interactions of children with ASD [22], and individual responses to emotional content such as nude imagery and faces [69]. Sensitive information, such as personality traits [36] and neurological diagnoses [47], could be linked to individuals that contributed to the aggregate data. To protect against this type of attack, differential privacy techniques have been proposed for securing heatmaps and other gaze-based features [50, 74]. However, they are constrained to dealing with already recorded gaze data and not real-time streams.

Mobile eye trackers rely on videos from an eye camera that captures the user's eye, and a front facing scene camera that records what they see. The scene camera is akin to wearable devices that are always on and recording video data. Public perception of these devices is overwhelmingly negative, as seen with the initial release of the Google Glass, as they infringe on the privacy of both the user and bystanders [20, 59, 70]. Daily users of eye tracking technology trade-off the privacy of their everyday actions for the benefit of activity logging, gaze-based interfaces, and assistive applications [7, 37, 55, 81]. Steil et al. have developed a privacy approach specifically for the scene camera, using a controlled shutter to disable the video feed in private situations [75]. The eye camera is unique in that it captures raw eye movements and personally identifying information without any layer of security. Previous approaches for wearable-based privacy and security do not apply to this context. This paper focuses on a solution to protect against unauthorized iris-based identification from eye images.

Iris Authentication Infrared images of the eye with sufficient resolution capture iris patterns unique to the individual. Iris recognition places in the top tier of biometrics as it is universal, distinct, permanent, and robust against spoofing attacks [39]. It is important to keep the iris pattern secure, as recognition methods are robust to poor lighting [44], off-axis imaging [16], occlusion [17], and distance [8], making the biometric accessible at times when the user may not consent. Iris authentication has been long established through the work of John Daugman [18] and many others[1], as a statistically valid method for recognition of an individual. As a result, iris patterns have been trusted for identification at voting booths [5], border customs [2], schools [1], and in hospitals [3]. These applications highlight the sensitivity of information that could be accessed if a hacker is able to steal identity through a biometric. Thus, the presence of a user's iris within a dataset or application places the user's identity at risk.

Defocus-based Identity Preservation Rana and colleagues presented a systems argument for why applications that process images and videos do not necessarily need access to the raw image feed [40, 41]. Neustaedter et al. [60] explored adding blur to increase privacy of a tele-conference video feed. They found that there is no general purpose blur level that preserves utility across all scenarios in this context. For example, the participants specified a much higher amount of blur in video that captured embarrassing activities such as picking their nose or changing clothes, compared to daily computer work. Participants were asked to identify the activities being performed in each video, with the level of blur being decreased until they could confidently classify the activity. The computed blur thresholds and classification rate determine that blur is effective at increasing privacy while retaining utility, but that the trade-off must be evaluated across applications and sensors. Hasan et al. [34] investigated various image filters such as masking, blurring, and pixelation with respect to their effectiveness in obscuring specific features of the content as well as retaining the utility and aesthetics of the photograph. They reported that blur was effective at obscuring the gender of the photographed person, though not so much the ethnicity or expression. Ultimately they determined that there is no 'one size fits all' solution for every scenario, and object size or security context can influence the optimal method. Pittaluga and Koppal [65] have implemented a similar blur-based privacy approach within the context of micro-scale image sensors. A hardware-based approach is used to add blur, as opposed to a software-based Gaussian blur. The use of optics to scatter light before the image is captured creates blur on the camera sensor. Applications like head tracking, person tracking, and facial recognition are explored with several types of camera sensors (thermal, IR , RGB) imaging the user. Each camera configuration and application must be optimized and designed to balance the trade-off between security and utility. Our work investigates adding blur to eye images pre-capture, however the goal is to do so without modifying the stock hardware or optics. This allows consumers to control their own privacy, as current consumer technology would lack any specialized privacy hardware.

John et al. [42] have proposed the only existing method to protect the iris biometric within eye tracking images. Gaussian blur is applied to the eye images to remove high frequency details from iris patterns. A monocular glasses-based eye tracker was used to collect data from five participants, in which eye images were captured and matched to each other. The authors found that an eye camera at 320x240 resolution is able to capture iris patterns that successful identify each individual without false positives. Their results suggest that a Gaussian blur with $\sigma = 5$ pixels is needed to reach the highest level of privacy, where no frames from any individual could be correctly recognized. Utility for the collected data was determined by an on-screen target viewing task, where blur at $\sigma = 5$ produced gaze error of less than 1.5° visual angle. At higher levels of blur pupil detection rates drastically decrease to 60% on average, resulting in a gaze data stream with gaps and low confidence values. Applying blur in software creates a risk that image data could be compromised before security is enforced. Our contribution to the state of the art is to explore the theoretical basis of a defocus-based solution, and propose and evaluate a hardware instantiation that is compatible with a popular eye tracker design. Importantly, we investigate the security-utility tradeoff of defocus parameters when utility is defined as how a virtual avatar is perceived, rather than data-level numerical error as in John et al. [42].

3 Security Vulnerability and Solution

The newest wave of VR and AR devices include integrated eye tracking devices, and are susceptible to identity theft and spoofing attacks. In

[1] Please see [27] for a review.

this section we describe the threat model that puts the user at risk, and propose defocus as a solution to enable secure eye tracking configurations. We provide a theoretical basis of a solution and evaluate it with respect to degrading the iris biometric and errors in gaze estimation while viewing on-screen targets.

3.1 Threat Model

Iris patterns present in eye tracking data streams can serve as a password, and are continuously streamed when an eye tracker is in use. This data stream is subject to a man-in-the-middle attack if images are sent over a network. In configurations where images are not streamed over a network, they are still subject to attacks when data is transferred at the hardware level.

An approach to protecting eye images is to only stream gaze data that is relevant to the application [6]. Image data is encapsulated within a processing unit, reducing the chance that a malicious user can gain access. However, this also restricts applications that may utilize the iris for improved gaze estimation [14], realistic rendering of the user's eye [26], and iris authentication in cases where it is desired, such as logging into the Microsoft Hololens 2.

Beyond risks in how manufacturers handle this sensitive data, the user must also control the permissions for third party applications that may access gaze and eye image data. There are growing concerns over how companies, large and small, handle sensitive data, with Facebook having their largest security breach most recently in 2018. A simple approach that alleviates all of these risks is to remove iris features prior to the image being recorded by the camera sensor. Then, even if a hacker gains access to the images they will not be able to use the iris patterns for authentication. Our work takes this approach.

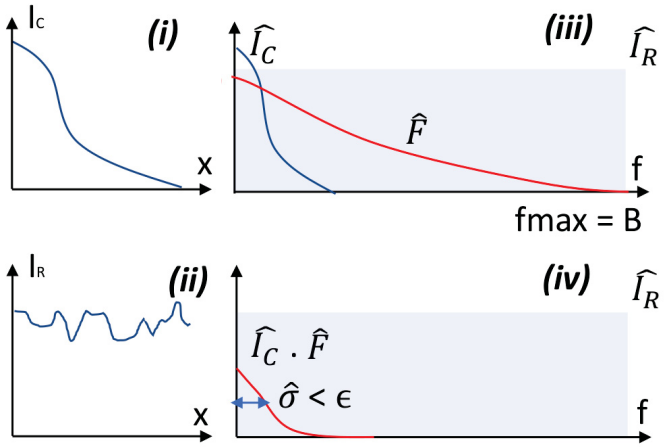

Fig. 1: In (i-ii) we depict, in 1D, the eye tracking image component I_C and the iris texture component I_R. In (iii) we show these again in the frequency domain, where the filter F is depicted as Gaussian blur and $\hat{I}_R$ is some distribution spanning the shaded region. In (iv) we see I_C remains usable for eye tracking when the standard deviation $\hat{\sigma}$ from $\hat{I}_C \cdot \hat{F}$ is not less than a minimum detectable bound ε.

3.2 Frequency-based argument for proposed solution

Our main assumption is that the eye tracking signal component I_C and the iris texture component I_R are *separable*, and the image can be expressed as $I = I_C + I_R$. Without loss of generality, we assume these are 1D functions, but all our arguments below hold for 2D signals. I_C is the component of the image that contains the eye tracking signal, and is modeled with a Gaussian distribution $I_C \approx N(\mu = 0, \sigma_C)$. This Gaussian disk captures eye tracking features, such as corneal reflection highlights or pupil extent [33]. The other component, I_R, corresponds to iris texture. Although iris textures have broad variation, we can assume that I_R is band-limited, since the highest frequency in the signal is limited by image resolution. Let us denote the largest possible frequency as B.

While I_C contains primarily low frequency content (defined by the standard deviation σ_C), I_R contains both low and high frequency content, with the higher frequencies being the identifying features (up to the maximum frequency B). Figure 1 illustrates the general functional form in spatial and frequency domain for these two signals.

Consider a low-pass filter F that is convolved with the image. We consider optical defocus, where the filter form is $F(x)$=$N(\mu$=$0, \sigma)$, i.e., a Gaussian blur. When I is convolved with $F(x)$, the result is

$$I_D(x) = I(x) * F(x) = I_C(x) * F(x) + I_R(x) * F(x) = I'_C(x) + I'_R(x). \quad (1)$$

Our claim: It is possible to select $F(x)$ such that the eye tracking features are still detectable in I'_C, while I'_R no longer contains the higher frequencies that enable iris-based authentication. Let $\hat{I}_C$ and $\hat{I}_R$ denote the Fourier transform of I_C and I_R respectively, and $\hat{F}$ denote the Fourier transform of F. Note that the Fourier transform of a Gaussian distribution is also a Gaussian with standard deviation $\hat{\sigma} = \frac{1}{\sigma}$, i.e., $\hat{F} \approx N(\mu = 0, \hat{\sigma}_F)$ and $\hat{I}_C \approx N(\mu = 0, \hat{\sigma}_C)$. In Fourier domain, Eq.1 is

$$\hat{I}_D(x) = \hat{I}_C(x) \cdot \hat{F}(x) + \hat{I}_R \cdot \hat{F}(x). \quad (2)$$

Upper bound (i.e. how much defocus is too much): If $\hat{\sigma}_F < \hat{\sigma}_C$, then the first term of Eq. 2 comprises the multiplication of two zero mean Gaussian functions, which yields a Gaussian function with zero mean and $\hat{\sigma} << \hat{\sigma}_C$. In the spatial domain, this corresponds to a Gaussian with $\sigma >> \sigma_C$. Intuitively, the corneal highlight or pupil extent has been heavily blurred, leading to difficulty in gaze location estimation. In frequency terms this means that the disk in the image is blurred enough that its transformed Gaussian dual has a indiscernible standard deviation, i.e. $\hat{\sigma} < \varepsilon$, where ε is vanishingly small. Thus, this imposes an upper limit on σ_F.

Lower bound (i.e. how much defocus is too little): Within the second term of Eq. 2, which contains the texture information necessary for iris-based authentication, the higher frequencies have been attenuated as a result of the point-wise multiplication with a Gaussian that has fallen off. Since $\hat{I}_R$ is band-limited by maximum frequency B, it's extent is within the range covered by $\hat{F}$. If σ_F is large enough the values of $\hat{F}$ are extremely small and "zero out" the values of $\hat{I}_R$ during point-wise multiplication. This imposes a lower limit on σ_F, such that the identifying features of the iris are removed from $I(x)$.

Our proposed approach: Optical defocus is produced by increasing the distance between the camera and the user's eye, forcing the iris region out of focus. While some configurations allow the camera to be adjusted or even feature a lens with adjustable focus, eye trackers with limited access to the camera may require additional optics or hardware to be installed. In our configuration the amount of defocus is controlled by varying the distance between the eye and the camera.

3.3 Implementation

We implement optical defocus to create a secure eye tracking configuration. We use a Pupil Labs Pro glasses-based eye tracker with an adjustable telescoping arm to increase camera distance. Example eye images from in-focus and out-of-focus configurations are shown in Figure 2. Eye trackers for XR devices use similar cameras, and this form of eye tracker is readily available to researchers and consumers. This is one instance of a secure eye tracking configuration. An example alternative configuration would be using an eye camera with an adjustable focus lens.

Camera Distance The out-of-focus configuration was implemented by increasing distance between the eye and camera to degrade iris authentication. First, the in-focus configuration was set up by placing the eye camera as close as possible to the user's eye, while keeping the eye in the center of the eye image frame. Then, to create the out-of-focus configuration the experimenter adjusted the telescoping arm to the farthest point, again orienting the camera such that the eye stayed within the frame. We compute the distance between the camera lens and the eye to quantify the impact of this process on gaze accuracy and iris authentication.

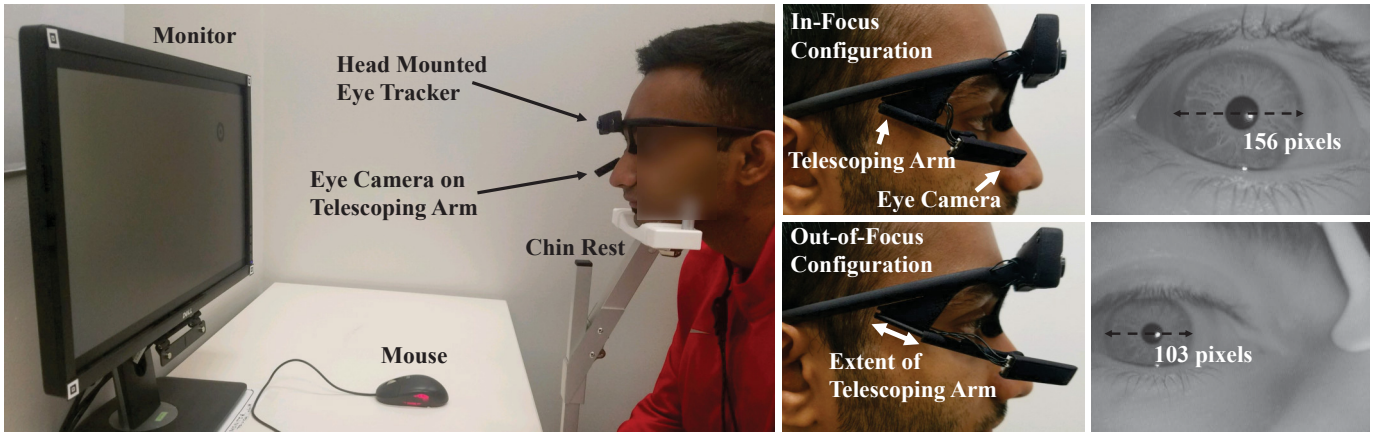

Fig. 2: Experimental setup for evaluation. The adjustable telescoping arm of the eye tracker is used to create an out-of-focus configuration. Eye images from an in-focus configuration (23.3mm) and out-of-focus configuration (35.4mm) are shown.

Camera distance is computed by modeling an imaging system with a thin lens. Figure 3 illustrates such a system. The distance between the iris and lens, and the lens and camera sensor are related by

$$\frac{W_{img}}{u} = \frac{W_{world}}{d}, \tag{3}$$

where d is the distance from the lens to the iris plane, W_{img} is the width of the iris as measured in the image, and W_{world} is the actual width of the iris. Variables d and u are related by the lens equation

$$\frac{1}{f} = \frac{1}{d} + \frac{1}{u}, \tag{4}$$

where f is the focal length of the camera in mm. We estimate W_{world} as the average width of a human iris, 11 mm [61], measure W_{img} within the image, and compute f.

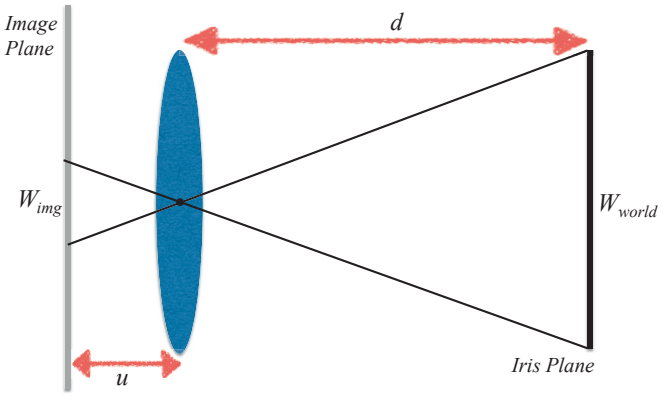

Fig. 3: Camera system imaging a flat iris plane with a thin lens. The lens equation produces the relationship $\frac{1}{f}=\frac{1}{d}+\frac{1}{u}$ and $\frac{W_{img}}{u}=\frac{W_{world}}{d}$.

As shown in Figure 3, the distance between the lens and camera sensor, u, is constant. The same process was followed to set up the in-focus eye tracker configuration for each participant. We compute u as follows: (1) For each participant Eq. 3 and Eq. 4 are used to solve for d, (2) the average distance for all participants, $\overline{d}$, is computed, (3) $\overline{d}$ is substituted into Eq. 4 to compute u. This process assumes that the measured values of d are within the depth of field of the camera for which the iris region is in-focus, and can be be estimated with the average distance, $\overline{d}$.

We computed the amount of defocus, σ, that was generated by an increase in distance from the in-focus configuration to the out-of-focus configuration. First, the out-of-focus distance d_{secure} is computed using $\frac{W_{img}}{u}=\frac{W_{world}}{d_{secure}}$ and W_{img} from an out-of-focus eye image. This is then substituted into Eq. 4 to generate a new u value, u', that represents the depth of a focal plane given the new camera distance. The amount of defocus σ in mm is then computed by

$$\frac{u'}{D} = \frac{u-u'}{\sigma}, \tag{5}$$

where D is the lens diameter measured to be 1.05mm, and σ represents the spread of a point that was in focus at the near distance, projected onto the camera sensor at the distance d_{secure}. σ is then converted from mm to pixels using the factor $0.003\frac{mm}{pixel}$, as specified in the OV9712-1D sensor spec sheet.

Iris Authentication For our experiment, iris segmentation was performed using an open source implementation of IrisSeg [29]. Our authentication procedure applies a bank of 1D Log-Gabor filters to the resulting iris pattern to generate a binary code that captures the identifying features of the iris pattern [44, 54].

To perform authentication the bit values of these codes are compared using Hamming Distance to determine if the source and target match. Hamming distance is defined as the number of bits that disagree between source and target binary codes,

$$\text{HD} = \frac{\|(\text{S}_{\text{code}} \otimes \text{T}_{\text{code}}) \cap (\text{S}_{\text{mask}} \cap \text{T}_{\text{mask}})\|}{\|\text{S}_{\text{mask}} \cap \text{T}_{\text{mask}}\|}, \tag{6}$$

where S_{code} and T_{code} are the input binary codes with their respective masks. The binary masks indicate which pixels contain the iris pattern, with zeros indicating eye lids, eye lashes, or any other detected noise [17]. In the subsequent data analysis iris codes are excluded if at least 75% of the bits are considered noise. We use a threshold of $HD_{auth} = 0.37$ to authenticate a match between source and target.

3.4 Evaluation

An iris authentication procedure is used to evaluate the increase in security from an in-focus configuration to out-of-focus configuration. Utility is measured using a target viewing task in a typical eye tracking setup, with error calculated between the estimated gaze positions and on-screen targets. Ideally, a secure configuration will degrade iris authentication while preserving the accuracy of gaze estimation.

Metrics The ability to authenticate a user is measured using the Correct Recognition Rate (CRR) [56]. CRR is computed as the percentage of frames between the source and target inputs where $HD < HD_{auth}$. Through this metric we determine that images collected during our "stop-and-stare" authentication routine can be used to identify the individual, with minimal false positives.

Eye tracking utility was measured in terms of gaze accuracy during the five target viewing task. The Pupil Labs software was used to identify frames with circular targets. These frames were used to calibrate a gaze mapping model that predicts the 2D gaze point-of-regard within

ID	In-Focus Distance (mm)	Out-of-focus Distance (mm)	Defocus σ (pixels)	In-Focus CRR (%)	Out-of-focus CRR (%)	In-Focus Gaze Error (°)	Out-of-focus Gaze Error (°)	In-Focus Precision (°)	Out-of-focus Precision (°)
S01	24.9	32.4	2.5	91	16	1.0	1.0	0.1	0.1
S02	30.0	35.8	3.6	87	3	1.4	1.3	0.1	0.1
S03	24.3	32.6	2.5	90	5	1.2	2.4	0.1	0.1
S04	24.6	31.7	2.2	95	15	2.2	2.0	0.1	0.1
S05	34.3	37.1	3.9	42	0	1.5	1.8	0.2	0.1
S06	26.1	38.8	4.4	73	0	1.7	1.1	0.1	0.1
S07	26.3	33.4	2.8	82	21	1.4	2.7	0.1	0.1
S08	29.4	36.3	3.7	78	1	1.4	1.6	0.1	0.1
S09	24.6	37.6	4.1	66	1	1.8	1.4	0.2	0.2
S10	24.0	34.7	3.2	86	0	1.1	1.4	0.1	0.1
S11	27.5	37.8	4.1	99	0	0.8	1.5	0.1	0.2
S12	25.4	33.7	2.9	78	27	2.0	2.1	0.1	0.1
S13	26.5	33.9	3.0	76	17	1.9	1.5	0.1	0.2
S14	24.9	37.4	4.0	79	0	0.9	1.6	0.1	0.1
S15	30.0	31.8	2.2	57	0	0.9	1.9	0.1	0.1
Mean	25.1	33.1	3.3	79	7	1.4	1.7	0.1	0.1
Std. Dev.	2.8	2.3	0.7	15	9	0.4	0.5	0.04	0.04

Table 1: Security and utility results for in-focus and out-of-focus eye tracking configurations. On average there was a difference of 8mm between in-focus and out-of-focus configurations. Defocusing the camera caused a decrease in CRR without an appreciable impact on gaze accuracy.

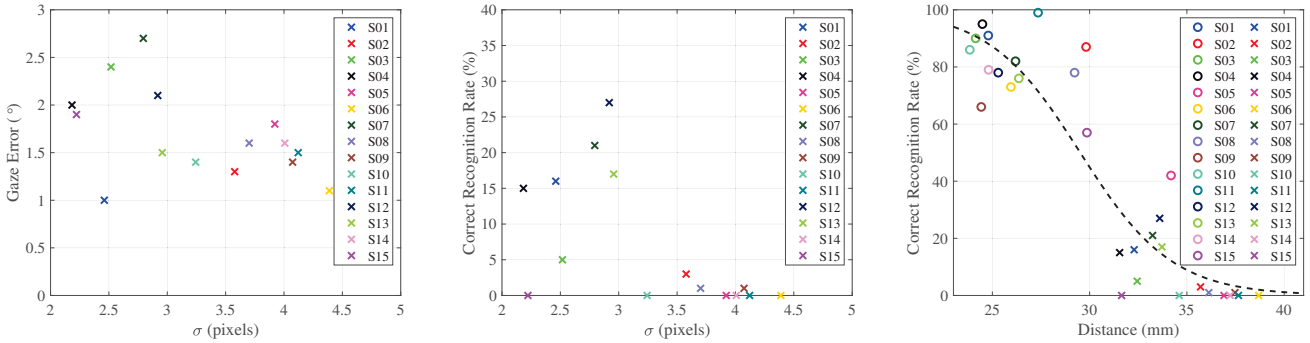

Fig. 4: Results for security and utility show that CRR is degraded by defocus (σ) and increased camera distance. Circles indicate data from the in-focus configuration, while crosses indicate data from the out-of-focus configuration. The dashed line represents a sigmoid curve fit to CRR as a function of distance. Angular error measured between targets and gaze data for the out-of-focus configuration was at most 2.7°.

the scene camera feed [45]. Once calibrated, the average error between projected gaze and the center of each target was computed in terms of visual angle. We computed the precision as the Root Mean Square Deviation between successive gaze locations while the targets were present to measure the stability of the gaze at each target.

Method We modify the experimental setup of John et al. [42] and evaluate the proposed hardware solution. Users sat with a chin rest and viewed circular targets presented on a desktop screen (see Figure 2). First, the eye tracker was set up for an in-focus configuration, as specified in Section 3.3. A video with five circular targets appearing for four seconds each was then shown, generated with Pupil Labs. The eye tracker was calibrated offline using Pupil Lab's default *3d_calibration* routine from Pupil Labs, with gaze samples and ground truth collected when the targets were present.

The user was asked to look directly at the eye tracking camera for five seconds, simulating a "stop-and-stare" interface for iris authentication [16, 66], prior to target viewing and directly afterwards. Only images from this part of the data collection were used for authentication, ensuring that the pupil and iris are on-axis with the camera. On-axis images increase the reliability of iris segmentation and matching [16, 80]. Each user logged around 300 frames during this procedure.

Eye tracking data was collected at 30 Hz using a Pupil Labs Pro glasses-based eye tracker (ca. 2016) with an eye image resolution of 320x240 [45]. We calibrated the fixed focus Pupil Labs eye camera using a checkerboard pattern and MATLAB's Single Camera Calibrator App to compute a focal length, f, of 338.04 pixels (1.014mm). Prior to analysis, frames containing blinks or motion blur were removed.

Participants Eye tracking data and images were collected from fifteen participants (8 male, 7 female) in an IRB approved user study. Participant demographics were 20% Asian, 13% Hispanic, 13% African American, 27% Indian, and 27% Caucasian.

Results We computed an authentication threshold, $HD_{auth} = 0.37$, based on the distributions of inter-class and intra-class HD values for our in-focus eye images. These distributions and HD_{auth} are illustrated in Figure 5. For our entire dataset, $HD_{auth} = 0.37$ creates an overall true positive rate of 60.1%, a false negative rate of 39.9%, true negative rate of 99.9988%, and a false positive rate of 0.0012%. While a higher true positive rate may be ideal, increasing the value of HD_{auth} would also increase the the false positive rate and compromise the system.

The average Hamming Distance between participants i and j, $\overline{HD_{ij}}$, is shown in Figure 6, with white cells indicating $\overline{HD_{ij}}$ is less than HD_{auth}. Grey cells indicate that the source does not match the target. Our results generate no false positives when the authentication condition is $\overline{HD_{ij}} < HD_{auth}$.

Authentication by $\overline{HD_{ij}}$ has the highest accuracy when comparing in-focus images with in-focus images. As expected, in-focus images did not create any matches with out-of-focus images. However, out-of-focus images did create matches with other out-of-focus images from the same individual, albeit less frequently. Only three participants produced a $\overline{HD_{ij}}$ less than HD_{auth} for the out-of-focus images.

Table 1 reports the CRR values, with an average in-focus CRR of 78.6%, while the out-of-focus images had a rate of 7.1%. Fig-

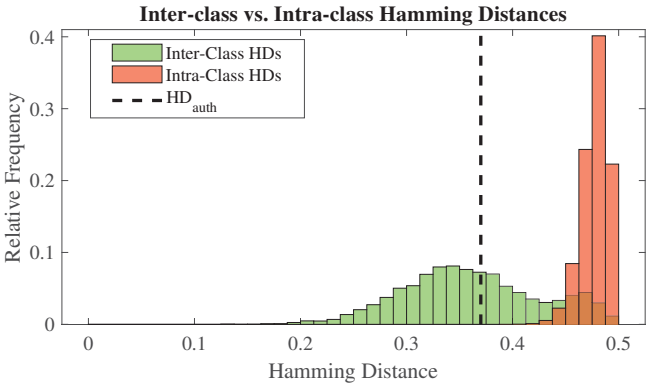

Fig. 5: A histogram of inter-class and intra-class Hamming Distances is used to determine a threshold for iris authentication.

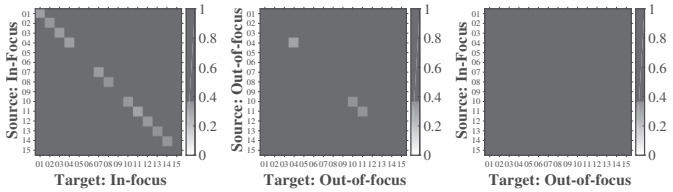

Fig. 6: Average Hamming Distance between source and target comparisons between each individual. Values less than the HD_{auth} indicate a match, and are colored white.

ure 4 (Right) demonstrates the relationship between camera distance and CRR by fitting a sigmoid function of the form $f(d) = \frac{1}{1+e^{-(a \cdot d+b)}}$, where $a = -0.43$, $b = 12.10$, and d is the input distance in mm. At 30mm CRR was 45%, and by 35mm CRR has dropped to 8%, showing that only a small percentage of frames can successfully authenticate the user at increased distances.

The computed σ values from Eq. 5 are presented in Table 1. Figure 4 (Left & Center) shows gaze accuracy and CRR respectively for the out-of-focus configuration as function of σ. Using σ to measure defocus allows us to compare results across configurations independent of the implementation, such as with software-based Gaussian blur.

Table 1 contains these average gaze error and precision values for each participant and configuration and Figure 4 (Left) demonstrates the relationship between out-of-focus gaze error and σ. We found that the average error across participants for the in-focus and out-of-focus configurations were 1.4° and 1.7° respectively. These values both fall within the magnitude for noise in an eye tracking system [35]. Ten participants saw an increase in error from the out-of-focus configuration, with the maximum error being 2.7°. Error within this range is acceptable for target viewing, where the targets span approximately 7°.

Discussion Using HD_{auth} we computed CRR for each participant, comparing every eye image with every other eye image. For the in-focus configuration we found on average 78.6% of frames were a match. This indicates that a login procedure that matches only one input image to another may not be robust enough for a consistent user experience. Using a larger HD_{auth} would create a smoother process, but compromise security. Instead, collecting a small set of ideal on-axis images and computing the average hamming distance from the reference may be a more dependable approach.

Previous research has shown that camera distance degrades iris authentication [42], based only on data from one user. We paramaterized and evaluated this approach extensively, and found that the average error of 1.7° introduced with this secure configuration did not have considerable impact on a target viewing task. While this error is acceptable for target viewing, it may not be for more complex applications such as animating the eyes of a social virtual avatar.

4 Evaluation of Utility for Avatar Gaze

Social virtual avatars have eye animation with the goal of increasing social presence and immersion. A large body of work has established that the animation of eye movements impacts viewer perceptions of avatar attributes, such as truthfulness and attentiveness [25, 72, 77]. The goal of this section is to determine how the noise introduced by secure eye tracking impacts the perception of animated virtual avatars.

4.1 Research Questions & Expected Outcomes

We conducted perceptual studies to answer the following questions:

- RQ_1: At what level of defocus do viewers detect a difference in the animation of a virtual avatar's eyes compared to a reference?
- RQ_2: What is the relationship between eye image defocus and the perception of avatar truthfulness, naturalness, attentiveness, comfort, and eye contact?

For RQ_1 we hypothesize that a medium amount of defocus, i.e., less than or equal to $\sigma = 3$ pixels, applied to the image feed will be detected by the viewer at a rate near chance, while data from larger values of σ will be detected at a higher rate. Past work has shown that the pupil detection rate declines after $\sigma = 3$ [42], which would result in a halt in eye animations during frames where the pupil was not detected. Additionally, the offset in gaze required for a viewer to indicate there is no longer mutual gaze varies across viewing distance and display mediums, ranging from less than 1° up to 9° [28]. Our results in Section 3.4 indicate gaze error for all values of σ up to 4.4 were more than 1° and less than 3°, falling within the range for mutual gaze with a virtual human face.

For RQ_2 we selected the attributes truthfulness, naturalness, attentiveness, comfort, and eye contact as they are influenced by avatar eye movements. We again hypothesize that up to a medium amount of defocus, i.e., less than or equal to $\sigma = 3$ pixels, there will be no difference in how eye movements are perceived by viewers. For defocus greater than $\sigma = 3$ we expect negative responses, or values less than 3 ('Neither Disagree or Agree') on the measured Likert scale.

4.2 Study 1: Detection Threshold

The goal of study 1 is to answer RQ_1. We designed a same-different experiment where naive viewers are presented with a reference avatar with unmodified eye tracking and a stimulus avatar with modified eye tracking, and they are tasked with reporting whether the two avatars are identical or different from each other. We compute psychometric curves from the participant responses and report the point of subjective equality (PSE) and detection threshold (DT). These values clarify the level of defocus at which viewers are able to perceive a difference in the eyes of the virtual avatar.

Stimuli Generation Naturalistic gaze data was recorded with the Pupil Labs Pro glasses-based eye tracker in a conversational scenario. We selected an English as a second language instructional video from YouTube[2]. The details of the video are shown in Table 2. The video had an instructor speak conversational sentences in English for the student to pause and repeat back to them. The topic of the conversation was a technique for learning English grammar. One of the authors watched the video, and acted out the part of the student by repeating sentences back as appropriate while being eye tracked. We extracted six 12 second segments from different parts of this dataset, resulting in six eye animations. Gaze directions from these segments were transferred on to a virtual avatar.

The virtual avatar was animated and rendered with the Unity game engine, version 2017.4.24f1. We created a model with Character Creator 3. We chose a bald male avatar with a realistic appearance to avoid simulating hair. Only the eyes of the avatar were animated; the rest of the face was static without any eye movement. The model was rigged to animate both eyes using monocular gaze data, as our eye tracker only records movements of the right eye. A reference gaze vector, $< x, y, z >$, was recorded with the author looking straight ahead at the beginning of data collection. This vector is used to generate a gaze offset vector, $< -x, -y, 0 >$, which when added to the reference gaze direction creates a 'forward' vector $< 0, 0, z >$. Using the 'forward'

[2] https://tinyurl.com/yxetvjw8

ID	Video Timestamp	Sentences spoken
1	0:30.73 - 0:42.73	"You may have been wondering, well, how am I going to learn to speak English properly if I don't study English grammar? Well, today I am going to talk to you about that. So, how is it that we do this?"
2	0:44.26 - 0:55.26	"There is a special technique and it is very easy. Research has shown that it is the best way to learn English grammar, or grammar for any language."
3	0:59.51 - 1:11.51	"It's called point of view stories, or point of view mini stories. A mini story is just a small story. So, how is that we use point of view stories to learn English grammar?"
4	1:20.10 - 1:32.10	"We listen to a number of different points of view for this very story. And by point of view I mean that we change something in the story, like the time that the story is being told."
5	1:47.58 - 1:59.58	"Let's start with an example. Now, when I usually teach these in a classroom, I'll start by telling it in the present tense. So let's start there."
6	2:28.18 - 2:40.18	"Make sure you understand it. So your next question may be, how are we going to use this story to learn English grammar? And that's a good question. As I said before, we are going to hear this story told in a number of different ways."

Table 2: Conversational sentences spoken and their timestamps in the video[2] while eye movements were recorded for each animation stimulus.

vector the avatar eye is oriented straight ahead towards the viewer. For each gaze vector in the animation data stream the gaze offset is added to generate a current gaze direction, which is then used to orient each eye. Gaze shifts relative to the gaze offset create the eye movements seen in the stimuli. The animations did not include any audio.

Fig. 7: Eye animations from blurred and unblurred images were presented side-by-side with identical avatars. The deviation in gaze from defocus ($\sigma = 5$) is shown in the avatar on the right.

For each of the six eye animations, we had the recorded eye images from which gaze positions are estimated. The original eye images were blurred using MATLAB's *imgaussfilt* function at five levels of σ, defined in pixels. The blurred images were fed through the Pupil Pro gaze detection pipeline, and the gaze positions estimated from the modified eye images were recorded. This procedure ensured that the internal parameters in the processing pipeline for the eye images were constant. It may be possible to tune the internal parameters in the processing pipeline to fit blurred input eye images. We left this tuning and its evaluation for a future experiment.

The Unity camera was positioned such that the rendered face spanned 8 visual degrees in our experimental setup, consistent with the size of a human face at a distance of 1.5m. This design choice was based on prior studies that use a distance of 1.5m or more when evaluating real face to face conversations and virtual interactions [11, 53]. Figure 7 shows the experiment setup with the side by side avatars. The reference and stimulus were not labeled. The left and right placement was swapped to avoid bias, resulting in a total of 60 trials per participant (6 eye animations $\times$ 2 positions (L/R) $\times$ 5 levels of σ).

Method We created a same-different task where each trial consists of a stimulus and a reference presented simultaneously, and participants are asked to indicate if they are the same or different [15, 23]. The Miss Rate is computed as the proportion of 'same' responses. The Point of Subjective Equality (PSE) corresponds to the 50% Miss Rate, i.e., the stimulus level at which participants are as likely to detect as they are to miss the difference between the stimulus and the reference. In other words, the PSE clarifies when a viewer can discriminate the presence of the stimuli at the same rate as chance. The Detection Threshold (DT) provides an upper bound on the amount of defocus that can be applied before a difference is perceived by a viewer.

We select the value of σ that corresponds to a 25% Miss Rate as the DT, where the participant is expected to consistently respond that the two animations are different. This threshold has been previously adopted in several virtual reality pyschometric experiments, as it is halfway between chance and a perfect detection rate [12, 13, 76, 85].

Our experiment uses a within-subjects design. Participants were presented with avatar eye animation generated from five levels of defocus in a randomized order. Both the stimuli and reference were shown on screen at the same time. Before starting the experiment participants were given the following prompt to describe the task: "In each trial you will be shown two videos of an animated virtual avatar side by side for 12 seconds. Only the eyes will be animated. Your task is to indicate whether the two animations presented are the same or different. Again, only the eyes are animated, so there will be no differences in other regions of the face. You will provide your response after each trial." A break was offered to participants halfway through the experiment. After completing all of the trials, participants filled out a post-study questionnaire that indicated their age, gender, ethnicity, and prior experience with virtual reality displays. The experiment took approximately 25 minutes. The experimental setup was as shown in Figure 2. Eye tracking data from a remote device mounted to the monitor was recorded, but is not discussed in our analysis.

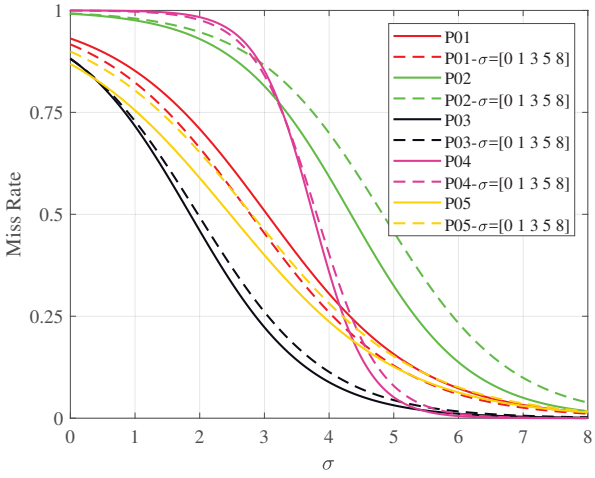

Fig. 8: Study 1 pilot results. Colored solid lines indicate psychometric functions fit to responses for all defocus levels, $\sigma = [0, 1, 2, 3, 4, 5, 6, 7, 8]$, for five participants and dashed lines indicate the corresponding functions fit to the subset $\sigma = [0, 1, 3, 5, 8]$.

Pilot experiment A five participant pilot experiment was conducted to establish the number of defocus levels (σ) to evaluate in Study 1. Nine levels, $\sigma = [0, 1, 2, 3, 4, 5, 6, 7, 8]$, were used. We found that the experimental duration exceeded 30 minutes and resulted in participants reporting fatigue even with mandatory breaks. Individual psychometric functions are shown in Figure 8. We then computed individual psychometric functions if only a subset of these nine levels were presented to the participant. We found that using only the defocus levels corresponding to $\sigma = [0, 1, 3, 5, 8]$ led to psychometric curves that were

comparable (varied by at most 0.6 σ). Reducing the number of defocus levels reduced the the experiment duration by approximately 11 minutes. Hence the subset was selected for the main experiment.

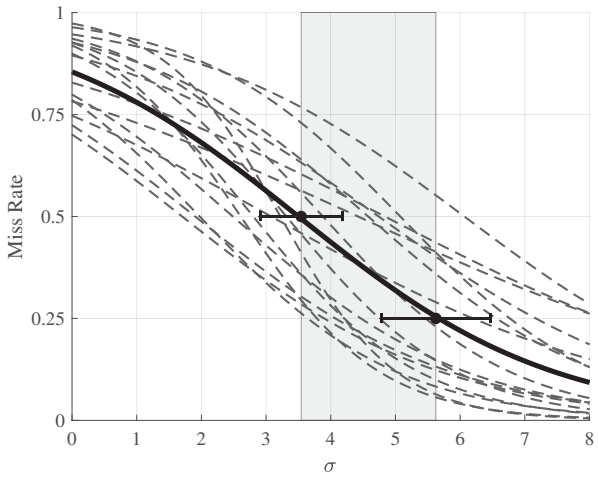

Fig. 9: Resulting psychometric functions of the same-different task for individual and pooled responses. Gray dashed lines represent individual responses, and the solid black line represents a function fit to the average responses across individuals. Error bars represent the 95% confidence interval for PSE and DT values.

Participants Twenty participants (11 male, 8 female, 1 "Preferred Not To Answer") with age ranging from 18 to 26 years were recruited from the university community under an IRB approved protocol. All participants reported normal or corrected-to-normal vision. Four participants reported no prior experience with a VR HMD. Of the participants that had experience with VR HMDs, the majority (75%) have used more than one type of HMD, such as the Oculus Rift DK2, Oculus Go, Google Cardboard, Samsung Gear VR, and HTC Vive.

Results The responses collected from each participant consisted of 60 categorical data items ('same'/'different'). Responses for each participant were grouped by level of defocus. Miss Rate was then calculated for each σ as the number of 'same' responses divided by the total number of responses. Miss Rate represents the probability that the participant does not detect a difference between animations.

Response quality was validated using the Miss Rate for stimuli where $\sigma = 0$. These stimuli showed two identical animations on screen, and if the participant responded 'different' 50% of the time or more, they either misunderstood the task or did not follow instructions. Four participants were removed from analysis using this criteria. A psychometric sigmoid function, defined as $f(\sigma) = \frac{1}{1+e^{-(a\sigma+b)}}$, models the probability that a participant would answer 'same' as a continuous function of σ. We fit a and b to each participant's responses using MATLAB's *glmfit* function. PSE and DT are then computed for each individual. A pooled psychometric function was computed by averaging Miss Rate across participants, and then fitting a function to the values, see Figure 9. The shaded region indicates a usable range of defocus for each individual curve, marking the area between the PSE and DT where a difference is not consistently perceived.

Discussion As shown in Table 3, the individual PSE values ranged from 1.70 to 6.06, and DT ranged from 3.64 to 8.38. The variation between individuals is also illustrated in Figure 9, as responses for Miss Rate ranges from 0.70 to 0.97 at $\sigma = 0$, where there is no difference between the two presented animations. Another source of this variation might stem from the strategy participants used to detect differences. Several participants indicated after the experiment that they examined the amount of visible sclera on either side of the avatar's iris in both animations to determine if there was a difference, while others indicated they relied on the movement during shifts in gaze direction. We expect participant responses to be more accurate during our experiment than a typical interaction with an avatar, as they are informed to look for differences from a reference presented side by side with the stimuli.

Participant	PSE	DT	Participant	PSE	DT
S01	3.22	4.21	S14	2.61	3.79
S02	1.90	4.07	S15	1.70	3.90
S05	4.58	6.56	S16	2.80	4.21
S06	4.42	8.21	S17	3.31	4.75
S07	6.06	8.38	S18	1.97	3.64
S08	4.74	7.17	S19	4.84	8.21
S09	3.08	6.20			
S12	2.49	4.47	**Average**	3.50	5.67
S13	3.88	5.46	**Std. Dev.**	1.30	1.73

Table 3: PSE and DT for each participant in Study 1.

Our analysis shows that the defocus value of $\sigma = 3.50$ is the average PSE, i.e., at this defocus level the viewer has as much chance of perceiving difference in eye animation relative to the original reference as she does to not perceive any difference. The average DT is $\sigma = 5.67$ pixels, which is the defocus level at which there is a 75% chance that viewers will be able to detect that the eye animation of the avatar is different compared to the original. These findings connect well with the results reported by John et al. [42]. They reported that defocus produced with $\sigma = 3$ degraded iris authentication for most individuals and $\sigma = 5$ completely degraded iris authentication.

The implications of our findings taken together with John et al.'s reports are as follows: if a user is comfortable with a moderate level of security, they can use up to $\sigma = 3.50$ of defocus without a noticeable effect on the eye animation of their social virtual avatar. Some users may want a higher level of security, and if they select a defocus of $\sigma = 5$ there is some chance the noise will be noticed in exchange for their preferred security level. It is likely though that when the reference animation is not shown it will be more difficult for a viewer to notice that the eye animation is modified. It is possible that in this case viewers may feel that "something is off" and report that the avatar did not make eye contact with them, or that the avatar did not pay attention to them, or that the avatar was not truthful or natural. We investigated these judgements in our next experiment.

4.3 Study 2: Avatar Attributes

The goal of this experiment is to answer RQ_2. We measure responses to truthfulness, naturalness, attentiveness, comfort, and eye contact for increased values of σ. Five levels of σ are considered: None ($\sigma = 0$), Low ($\sigma = 1$), Medium ($\sigma = 3$), High ($\sigma = 5$) and Very High ($\sigma = 8$).

Stimuli Generation The animation renders from Study 1 were used for Study 2. However, instead of two animations being presented only one animation was shown at a time in the center of the screen.

Method The study structure and apparatus was identical to Study 1, except participants rated each animation. The prompt provided to participants was: "In each trial you will be shown a video of an animated virtual avatar for 12 seconds. Only the eyes are animated. Imagine you are having a conversation with the avatar. Your task is to respond to several prompts about the animation after each trial.". Based on prior work we evaluate each interaction in terms of truthfulness [77], naturalness [43], attentiveness [25], comfort with the avatar [32], and eye contact [58]. After watching each animation the participant used a mouse to respond to the following prompts, using a five point Likert scale from 'Strongly Disagree' to 'Strongly Agree'(1-5):

(1) The avatar was truthful. (2) The eye movements of the avatar were natural. (3) The avatar paid attention to me. (4) I felt comfortable in the presence of this avatar. (5) The avatar made eye contact with me.

The experiment follows a within-subjects design, where every participant saw every animation and defocus level. Animations were presented in randomized order, and a break was offered halfway through the experiment. Each stimulus was presented and rated twice, leading to a total of 60 trials per participant (6 eye animations $\times$ 2 repetitions $\times$ 5 levels of σ). The experiment took approximately 35 minutes. Again, eye tracking data from a remote device mounted to the monitor was recorded and is not discussed in our analysis.

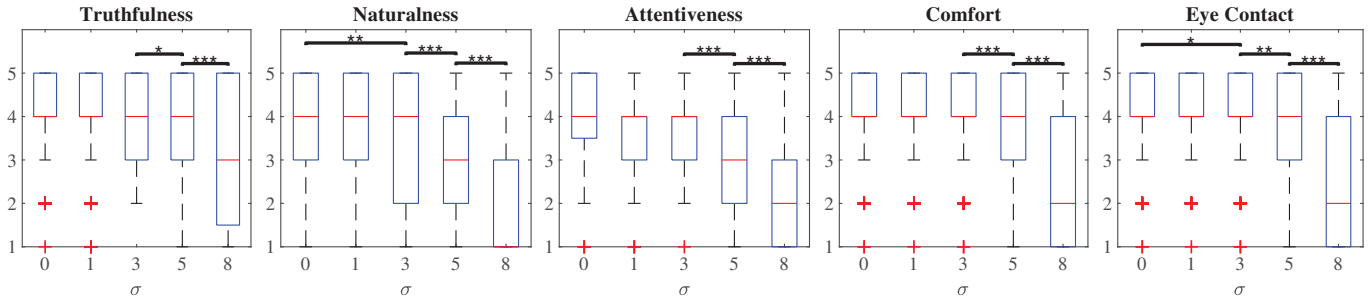

Fig. 10: Box plots indicating the median, 25%, and 75% quartiles for Study 2 results. Significantly different groups are marked with * when $p < .05$, ** when $p < .01$, and *** when $p < .001$. For clarity ** significance brackets were established but not drawn between groups σ=[0, 1] and σ=5, along with *** significance bars for groups σ=[0, 1, 3] and σ=8 across all attributes.

Pilot experiment A five participant pilot experiment was conducted to establish the number of repetitions for Study 2. We found that showing three repetitions of each stimulus resulted in participants asking for additional breaks after the halfway point, and the experiment could take up to an hour to complete. To limit the experiment duration and reduce participant fatigue we decided to use only two repetitions.

Participants Nineteen participants (14 male, 5 female) with age ranging from 19 to 39 were recruited from the university community under an IRB approved protocol. All participants reported normal or corrected-to-normal vision. Participants were ineligible if they had previously participated in Study 1, to ensure they had not previously seen the animation stimuli.

Results Likert scale responses for each dependent variable (truthfulness, naturalness, attentiveness, comfort, eye contact) represent ordinal data grouped by the defocus parameter σ. Figure 10 shows the average and standard error values for each attribute.

The Kolmogorov-Smirnov test for normality was applied to each group and variable. Data were not normally distributed ($p < 0.001$), and therefore non-parametric statistical tests were used. A Friedman test showed a significant main effect of σ for truthfulness ($\chi^2(4)$=162.72,$p < 0.001$), naturalness ($\chi^2(4)$=290.2,$p < 0.001$), attentiveness ($\chi^2(4)$=300.41,$p < 0.001$), comfort ($\chi^2(4)$=279.15,$p < 0.001$), and eye contact ($\chi^2(4)$=199.23,$p < 0.001$). For each attribute pairwise Wilcoxon signed rank tests with Bonferroni correction showed significant differences between $\sigma = 5$ and all other levels of σ ($p < .05$ or less); as well as between $\sigma = 8$ and all other levels of σ ($p < .001$). Additionally, for naturalness and eye contact significant differences were found between $\sigma = 0$ and $\sigma = 3$, with ($p < .01$) and ($p < .05$) respectively. Figure 10 visualizes the results as boxplots.

Discussion Our analysis shows that for $\sigma = 0$, i.e., no blur, average responses for each attribute were approximately 4, or 'Slightly Agree' on the Likert scale. Thus, participants agreed that the avatar was truthful, eye movements were natural, the avatar paid attention to them, they were comfortable with the avatar, and maintained eye contact. They did not 'Strongly Agree' with these statements, however. For truthfulness, naturalness, and comfort this is likely a result from only the avatar's eyes being animated and the lack of blinks. With respect to attentiveness and eye contact, the animations did not respond to the user's gaze, causing participants to provide only slight agreement. Still, at the end of the experiment several participants asked if the avatar was responding to their eye movements, as they knew they were being eye tracked. This indicates that the animation stimuli was convincing enough to simulate eye contact and interaction with the avatar.

Significance testing found a decrease in all response values at $\sigma = 5$ and $\sigma = 8$. Average responses for $\sigma = 5$ ranged from 3.04 to 3.78, indicating participants did not have a negative experience, but less positive. This is consistent with the findings from Study 1, as $\sigma = 5$ is near the DT ($\sigma = 5.67$). At $\sigma = 8$ only responses for truthfulness averaged to 3, i.e., 'Neither disagree or agree', which means the participants were not able to consistently determine if the avatar acted truthfully based on eye movements. Avatars using $\sigma = 8$ may be limited in their ability to immerse the viewer within a conversational setting. Averages for the rest of the attributes fell between 1.93 and 2.58, indicating that the avatar no longer convinced them. Eye tracking in the presence of this much defocus is not feasible for a convincing social avatar.

5 Conclusion

We have implemented and evaluated a novel hardware-based eye tracking configuration to secure the iris biometric from unauthorized identification. The secure configuration produced an average Correct Recognition Rate of 7% compared to 79% before defocus is introduced. Our second contribution is a pyschophysical experiment that determines the detection threshold for users viewing the eye movements of a virtual avatar animated using eye tracking data. Our results suggest that a defocus parameter of $\sigma = 3.5$ should be used if utility is preferred over security, and $\sigma = 5$ if security is preferred. Our third contribution is measuring the effect of σ on several attributes important for social interactions with virtual avatars, such as eye contact and naturalness. Results indicate attributes are degraded at σ values of 5 and 8, and the avatar no longer maintains eye contact, attentiveness, or naturalness.

Limitations The stimuli used for our perceptual evaluation has limitations. Particularly, our evaluation does not consider the impact of defocus on eye movement characteristics such as the blinks, the dynamics of saccades with large amplitudes, or estimated pupil diameter. These characteristics play an important role in complex social interactions and are more prominent the closer the user is to the avatar. The stimuli also did not include head or mouth movements. The defocus solution presented in this paper leveraged the telescoping arm of a popular eye tracker. More generally, a defocus solution applies to configurations where the eye camera is readily accessible, though future work might investigate clip on optics similar to Pittaluga and Koppal [65]. Our findings with respect to the fall in correct recognition rate are based on the Daugman method of iris recognition. If the iris recognition module were to be replaced with upcoming deep network based approaches, such as one proposed by Proenca and Neves [67], the fall in correct recognition rate as a function of hardware parameters might need to be re-assessed. Our work provides a foundation for developing an automated system that continually optimizes the security-utility trade-off even as new methods of eye tracking and iris recognition are invented.

Future Work It would be interesting to investigate an optimization framework for security and utility. It would also be useful to create implementations of secure eye tracking configurations that apply to different camera form factors. Additional perceptual experiments with a smaller distance from an avatar, and more realistic features on the avatar that include blinks, eyelid movements, and pupil diameter would provide further insight as well as implementing a similar evaluation within an immersive VR environment. Our work motivates active research in these directions *before* eye tracking in XR becomes ubiquitous and users are at risk to malicious attacks.

Acknowledgments

The authors wish to thank Cody LaFlamme for generating the stimuli used in our studies. Authors acknowledge funding from the National Science Foundation (Awards IIS-1566481, IIS-1514154, and IIS-1423189), and the NSF Graduate Research Fellowship (Awards DGE-1315138 and DGE-1842473).

REFERENCES

[1] Biometrics for education. https://www.iritech.com/biometric-education-Kenya. Accessed: 2019-08-22.

[2] Keeping an eye on security. https://www.schiphol.nl/en/page/how-the-iris-scan-works/. Accessed: 2019-08-22.

[3] Louisiana hospital adopts iris-based patient id system. https://findbiometrics.com/la-hospital-adopts-iris-based-patient-id-system-26203/. Accessed: 2019-08-22.

[4] Microsoft unveils hololens 2: Twice the field of view, eye tracking. https://arstechnica.com/gadgets/2019/02/microsoft-unveils-hololens-2-twice-the-field-of-view-eye-tracking/. Accessed: 2019-08-29.

[5] Somaliland election saw iris id technology deployed. https://www.biometricupdate.com/201801/somaliland-election-saw-iris-id-technology-deployed. Accessed: 2019-08-22.

[6] Tobii privacy policy. https://www.tobii.com/group/privacy-policy/. Accessed: 2019-08-29.

[7] T. Ahmed, P. Shaffer, K. Connelly, D. Crandall, and A. Kapadia. Addressing physical safety, security, and privacy for people with visual impairments. In *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*, pp. 341–354, 2016.

[8] A. A. Algawhari and Y. Huang. Iris recognition under unconstrained conditions. In *2018 International Conference on Image and Video Processing, and Artificial Intelligence*, vol. 10836, p. 108360D. International Society for Optics and Photonics, 2018.

[9] B. Bastani, E. Turner, C. Vieri, H. Jiang, B. Funt, and N. Balram. Foveated pipeline for AR/VR head-mounted displays. *Information Display*, 33(6):14–35, 2017.

[10] R. Bednarik, H. Vrzakova, and M. Hradis. What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In *Proceedings of the symposium on eye tracking research and applications*, pp. 83–90. ACM, 2012.

[11] J. K. Bennett, S. Sridharan, B. John, and R. Bailey. Looking at faces: autonomous perspective invariant facial gaze analysis. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 105–112. ACM, 2016.

[12] L. Bölling, N. Stein, F. Steinicke, and M. Lappe. Shrinking circles: Adaptation to increased curvature gain in redirected walking. *IEEE transactions on visualization and computer graphics*, 25(5):2032–2039, 2019.

[13] B. Bolte and M. Lappe. Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression. *IEEE transactions on visualization and computer graphics*, 21(4):545–552, 2015.

[14] A. Chaudhary and J. Pelz. Motion tracking of iris features to detect small eye movements. *Journal of Eye Movement Research*, 12(6), Apr. 2019. doi: 10.16910/jemr.12.6.4

[15] M. Cheetham, P. Suter, and L. Jäncke. The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Frontiers in human neuroscience*, 5:126, 2011.

[16] J. Daugman. New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1167–1175, 2007.

[17] J. Daugman. How iris recognition works. In *The essential guide to image processing*, pp. 715–739. Elsevier, 2009.

[18] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE transactions on pattern analysis and machine intelligence*, 15(11):1148–1161, 1993.

[19] E. J. David, J. Gutiérrez, A. Coutrot, M. P. Da Silva, and P. L. Callet. A dataset of head and eye movements for 360 videos. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pp. 432–437. ACM, 2018.

[20] T. Denning, Z. Dehlawi, and T. Kohno. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 2377–2386. ACM, 2014.

[21] Y. Dong and D. L. Woodard. Eyebrow shape-based features for biometric recognition and gender classification: A feasibility study. In *2011 International Joint Conference on Biometrics (IJCB)*, pp. 1–8. IEEE, 2011.

[22] H. Duan, G. Zhai, X. Min, Z. Che, Y. Fang, X. Yang, J. Gutiérrez, and P. L. Callet. A dataset of eye movements for the children with autism spectrum disorder. In *Proceedings of the 10th ACM Multimedia Systems Conference*, pp. 255–260. ACM, 2019.

[23] A. Duchowski, S. Jörg, A. Lawson, T. Bolte, L. Świrski, and K. Krejtz. Eye movement synthesis with 1/f pink noise. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, pp. 47–56. ACM, 2015.

[24] X. Feng, V. Swaminathan, and S. Wei. Viewport prediction for live 360-degree mobile video streaming using user-content hybrid motion tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2):43, 2019.

[25] Y. Ferstl, E. Kokkinara, and R. McDonnell. Facial features of non-player creatures can influence moral decisions in video games. *ACM Transactions on Applied Perception (TAP)*, 15(1):4, 2017.

[26] G. Francois, P. Gautron, G. Breton, and K. Bouatouch. Image-based modeling of the human eye. *IEEE Transactions on Visualization and Computer graphics*, 15(5):815–827, 2009.

[27] A. G. Gale and S. Salankar. A review on advance methods of feature extraction in iris recognition system. *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) e-ISSN*, pp. 2278–1676, 2014.

[28] M. Gamer and H. Hecht. Are you looking at me? measuring the cone of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3):705, 2007.

[29] A. Gangwar, A. Joshi, A. Singh, F. Alonso-Fernandez, and J. Bigun. Irisseg: A fast and robust iris segmentation framework for non-ideal iris images. In *Biometrics (ICB), 2016 International Conference on*, pp. 1–8. IEEE, 2016.

[30] M. Garau, M. Slater, S. Bee, and M. A. Sasse. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 309–316. ACM, 2001.

[31] S. Grogorick, G. Albuquerque, and M. Maqnor. Gaze guidance in immersive environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 563–564, March 2018. doi: 10.1109/VR.2018.8446215

[32] R. E. Guadagno, K. R. Swinth, and J. Blascovich. Social evaluations of embodied agents and avatars. *Computers in Human Behavior*, 27(6):2380–2385, 2011.

[33] D. W. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE TPAMI*, 32(3):478–500, 2009.

[34] R. Hasan, E. Hassan, Y. Li, K. Caine, D. J. Crandall, R. Hoyle, and A. Kapadia. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 47. ACM, 2018.

[35] K. Holmqvist, M. Nyström, and F. Mulvey. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*, pp. 45–52. ACM, 2012.

[36] S. Hoppe, T. Loetscher, S. A. Morey, and A. Bulling. Eye movements during everyday behavior predict personality traits. *Frontiers in human neuroscience*, 12:105, 2018.

[37] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia. Privacy behaviors of lifeloggers using wearable cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 571–582. ACM, 2014.

[38] Z. Hu, C. Zhang, S. Li, G. Wang, and D. Manocha. SGaze: A data-driven eye-head coordination model for realtime gaze prediction. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2002–2010, 2019.

[39] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[40] S. Jana, D. Molnar, A. Moshchuk, A. Dunn, B. Livshits, H. J. Wang, and E. Ofek. Enabling fine-grained permissions for augmented reality applications with recognizers. In *Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13)*, pp. 415–430, 2013.

[41] S. Jana, A. Narayanan, and V. Shmatikov. A scanner darkly: Protecting user privacy from perceptual applications. In *2013 IEEE symposium on security and privacy*, pp. 349–363. IEEE, 2013.

[42] B. John, S. Koppal, and E. Jain. EyeVEIL: degrading iris authentication in eye tracking headsets. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, p. 37. ACM, 2019.

[43] S. Jörg, A. Duchowski, K. Krejtz, and A. Niedzielska. Perceptual adjustment of eyeball rotation and pupil size jitter for virtual characters. *ACM Transactions on Applied Perception (TAP)*, 15(4):24, 2018.

[44] A. Kahlil and F. Abou-Chadi. Generation of iris codes using 1d log-gabor filter. In *The 2010 International Conference on Computer Engineering & Systems*, pp. 329–336. IEEE, 2010.

[45] M. Kassner, W. Patera, and A. Bulling. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and*

ubiquitous computing: Adjunct publication, pp. 1151–1160. ACM, 2014.

[46] E. Langbehn, F. Steinicke, M. Lappe, G. F. Welch, and G. Bruder. In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Transactions on Graphics (TOG)*, 37(4):66, 2018.

[47] R. J. Leigh and D. S. Zee. *The neurology of eye movements*. Oxford USA Press, 2015.

[48] T. Li, Q. Liu, and X. Zhou. Ultra-low power gaze tracking for virtual reality. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, p. 25. ACM, 2017.

[49] D. J. Liebling and S. Preibusch. Privacy considerations for a pervasive eye tracking world. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1169–1177. ACM, 2014.

[50] A. Liu, L. Xia, A. Duchowski, R. Bailey, K. Holmqvist, and E. Jain. Differential privacy for eye-tracking data. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, p. 28. ACM, 2019.

[51] S. Lombardi, J. Saragih, T. Simon, and Y. Sheikh. Deep appearance models for face rendering. *ACM Transactions on Graphics (TOG)*, 37(4):68, 2018.

[52] P. Lungaro, R. Sjöberg, A. J. F. Valero, A. Mittal, and K. Tollmar. Gaze-aware streaming solutions for the next generation of mobile VR experiences. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1535–1544, 2018.

[53] A. MacQuarrie and A. Steed. Perception of volumetric characters' eye-gaze direction in head-mounted displays. In *Proceedings of 2019 IEEE Virtual Reality (VR)*, vol. 2019. IEEE, 2019.

[54] L. Masek and P. Kovesi. Matlab source code for a biometric identification system based on iris patterns, the university of western australia, 2003, 2013.

[55] C. Mei, B. T. Zahed, L. Mason, and J. Ouarles. Towards joint attention training for children with ASD-a VR game approach and eye gaze exploration. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 289–296. IEEE, 2018.

[56] D. M. Monro, S. Rakshit, and D. Zhang. Dct-based iris recognition. *IEEE TPAMI*, (4):586–595, 2007.

[57] J. H. Mueller, P. Voglreiter, M. Dokter, T. Neff, M. Makar, M. Steinberger, and D. Schmalstieg. Shading atlas streaming. In *SIGGRAPH Asia 2018 Technical Papers*, p. 199. ACM, 2018.

[58] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 61–68. ACM, 2009.

[59] P. E. Naeini, S. Bhagavatula, H. Habib, M. Degeling, L. Bauer, L. F. Cranor, and N. Sadeh. Privacy expectations and preferences in an IoT world. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, pp. 399–412, 2017.

[60] C. Neustaedter, S. Greenberg, and M. Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(1):1–36, 2006.

[61] K. Nishino and S. K. Nayar. Corneal imaging system: Environment from eyes. *International Journal of Computer Vision*, 70(1):23–40, 2006.

[62] J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan, and H. Devos. Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1302–1311, 2017.

[63] Y. S. Pai, B. I. Outram, B. Tag, M. Isogai, D. Ochi, and K. Kunze. Gazesphere: Navigating 360-degree-video environments in VR using head rotation and eye gaze. In *ACM SIGGRAPH 2017 Posters*, p. 23. ACM, 2017.

[64] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)*, 35(6):179, 2016.

[65] F. Pittaluga and S. J. Koppal. Privacy preserving optics for miniature vision sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 314–324, 2015.

[66] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre. The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE TPAMI*, 32(8):1529–1535, 2009.

[67] H. Proenca and J. C. Neves. Segmentation-less and non-holistic deep-learning frameworks for iris recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0, 2019.

[68] V. Rajanna and J. P. Hansen. Gaze typing in virtual reality: impact of keyboard design, selection method, and motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, p. 15. ACM, 2018.

[69] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli, and T.-S. Chua. An eye fixation database for saliency detection in images. In *European Conference on Computer Vision*, pp. 30–43. Springer, 2010.

[70] Y. Rashidi, T. Ahmed, F. Patel, E. Fath, A. Kapadia, C. Nippert-Eng, and N. M. Su. "You don't want to be the next meme": College students' workarounds to manage privacy in the era of pervasive photography. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pp. 143–157, 2018.

[71] K. Ruhland, S. Andrist, J. Badler, C. Peters, N. Badler, M. Gleicher, B. Mutlu, and R. Mcdonnell. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Eurographics state-of-the-art report*, pp. 69–91, 2014.

[72] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. In *Computer graphics forum*, vol. 34, pp. 299–326. Wiley Online Library, 2015.

[73] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. Saliency in VR: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642, 2018.

[74] J. Steil, I. Hagestedt, M. X. Huang, and A. Bulling. Privacy-aware eye tracking using differential privacy. In *11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[75] J. Steil, M. Koelle, W. Heuten, S. Boll, and A. Bulling. Privaceye: privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, p. 26. ACM, 2019.

[76] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE transactions on visualization and computer graphics*, 16(1):17–27, 2009.

[77] W. Steptoe, A. Steed, A. Rovira, and J. Rae. Lie tracking: social presence, truth and deception in avatar-mediated telecommunication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1039–1048. ACM, 2010.

[78] Q. Sun, F.-C. Huang, J. Kim, L.-Y. Wei, D. Luebke, and A. Kaufman. Perceptually-guided foveation for light field displays. *ACM Transactions on Graphics (TOG)*, 36(6):192, 2017.

[79] Q. Sun, A. Patney, L.-Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. Mcguire, D. Luebke, and A. Kaufman. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Transactions on Graphics (TOG)*, 37(4):67, 2018.

[80] J. Thompson, H. Santos-Villalobos, M. Karakaya, D. Barstow, D. Bolme, and C. Boehnen. Off-angle iris correction using a biological model. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–8. IEEE, 2013.

[81] P. Venuprasad, T. Dobhal, A. Paul, T. N. Nguyen, A. Gilman, P. Cosman, and L. Chukoskie. Characterizing joint attention behavior during real world interactions using automated object and gaze detection. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, p. 21. ACM, 2019.

[82] M. Volonte, A. Robb, A. T. Duchowski, and S. V. Babu. Empirical evaluation of virtual human conversational and affective animations on visual attention in inter-personal simulations. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 25–32. IEEE, 2018.

[83] E. Whitmire, L. Trutoiu, R. Cavin, D. Perek, B. Scally, J. Phillips, and S. Patel. Eyecontact: scleral coil eye tracking for virtual reality. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, pp. 184–191. ACM, 2016.

[84] R. Zemblys and O. Komogortsev. Developing photo-sensor oculography (PS-OG) system for virtual reality headsets. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, p. 83. ACM, 2018.

[85] A. Zenner and A. Krüqer. Estimating detection thresholds for desktop-scale hand redirection in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 47–55. IEEE, 2019.

[86] G. Zhang and J. P. Hansen. Accessible control of telepresence robots based on eye tracking. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, p. 50. ACM, 2019.